Week 3 – Visualize Sunflower Data

YOUR NAME HERE

September 07, 2016

Quantifying and visualizing potential fitness

Last week each individual in class collected data on the geographic location, size (diameter at a height of 12"), and estimated number of seeds for individual sunflower plants. Stem diameter is a measure of plant size. The estimated number of seeds per individual is a measure of the potential fitness of a plant.

This week your job is to construct a data set that combines all the data from all individuals (in EXCEL) and uses the data to visualize the variation in a population.

The compiled data from each lab section are available on D2L.

Goal summary

What you should be able to do	**How it will be achieved**	**Questions to answer**
Assemble a large data set without error	Combine several small spreadsheet into one large spreadsheet using EXCEL	
Show the geographic distribution of the potential fitness of individuals	Make a map of the plants with circle size proportional to log(# seeds) using R	Is there any pattern to the geographic distribution of variation? Explain you answer.
Compare diameter at 12" with estimated number of seeds for all plants (bivariate plot)	Construct a script in R	What is the association between the diameter of the stems and the estimated number of seeds?
Compare two histograms for number of seeds per plant: one the untransformed data and one with the data log-transformed	Construct a script in R	What is different about the two ways of summarizing the data?
Use the mean number of seeds per plant as an estimate of the intrinsic rate of increase and, assuming the number of plants observed this year estimates population size, estimate the population size next year (assume all seeds survive)	Construct a script in R	What is the expected number of plants next year if all seeds survive?
Assume that population size next year is the same as this year and calculate the proportion of the total estimated reproductive effort of the population that will ultimately die (fail to germinate and grow)	Construct a script in R	Does the estimate of the fraction of seeds that will probably perish indicate that there is a struggle for existence? Why or why not?
Use one of the graphs you made a make a claim based on the evidence in the graph	Practice scientific thinking	

R Code

Question 1

Use the following code block to load the full sunflower data set and *answer the question directly below the code block.* NOTE: You *CANNOT* load the file using `file.choose()` function. Your document will not knit if you try. Instead uou *must* save this file in the same place as the data set, or use `setwd()` to direct `R` to the location of the data file.

```
#load in the sunflower data
flowers <- read.csv("all_sunflowerv2.csv")

#look at the structure of the data file
str(flowers)
```

```
## 'data.frame':    256 obs. of  7 variables:
##  $ name     : Factor w/ 112 levels "Abdul Samsudin",..: 41 108 86 57 15 74 74 111 99 44 ...
##  $ section  : int  24 1 2 2 14 2 2 2 13 2 ...
##  $ latitude : num  40 40 40 40 40 ...
##  $ longitude: num  -106 -106 -105 -105 -105 ...
##  $ base     : num  33 9 22 38 15 10 14 67 40 40 ...
##  $ heads    : int  8 10 59 28 6 3 5 113 54 58 ...
##  $ seeds    : int  299 145 245 134 105 54 136 297 140 189 ...
```

Question 2

Use the following code block to show the geographic distribution of the potential fitness of individuals by making a map of the plants with circle size proportional to log(# seeds). *Hint:* you'll need to add cex= in your call to `plot()`

```
#plot the data in a space defined by latitude and longitude and make point size proportional to the log
plot(flowers$longitude, flowers$latitude, cex=1, ylim=c(39.5,40.1), xlim=c(-105.3,-105.1), ylab="Latitu
```

Is there any pattern to the geographic distribution of variation? Explain you answer.

Your answer here – keep the asterisks

Question 3

Use the following code block to compare diameter at 12" with estimated number of seeds for all plants (bivariate plot)

```
#your code here
```

What is the association between the diameter of the stems and the estimated number of seeds?

Your answer here – keep the asterisks

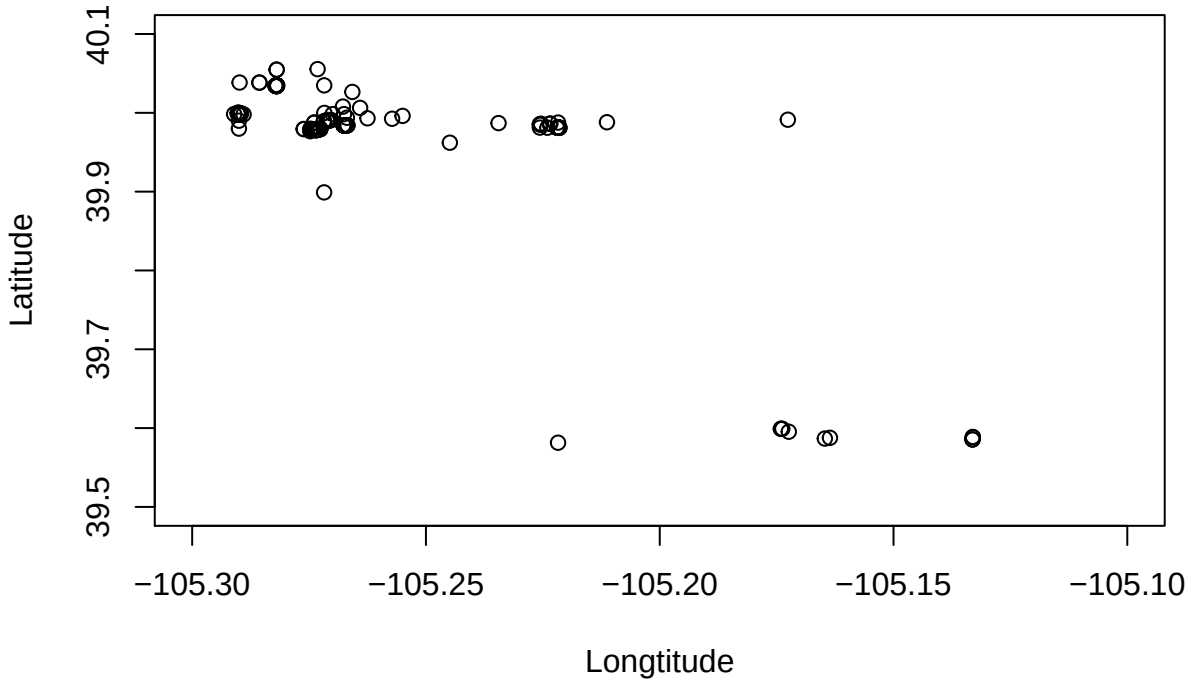

Figure 1: Geographic distribution of sunflowers with point size scaled to the log10 of the number of flowers

Question 4

Compare two histograms for number of seeds per plant: one the untransformed data and one with the data log-transformed &

```
#your code here
```

What is different about the two ways of summarizing the data?

Your answer here – keep the asterisks

Question 5

Use the mean number of seeds per plant as an estimate of the intrinsic rate of increase and, assuming the number of plants observed this year estimates population size, estimate the population size next year (assume all seeds survive)

```
#your code here
```

What is the expected number of plants next year if all seeds survive?

Your answer here – keep the asterisks

Question 6

Assume that population size next year is the same as this year and calculate the proportion of the total estimated reproductive effort of the population that will ultimately die (fail to germinate and grow)

```
#your code here
```

Does the estimate of the fraction of seeds that will probably perish indicate that there is a struggle for existence? Why or why not?

Question 7

Use one of the graphs you made a make a claim based on the evidence in the graph. Make sure to mention which figure you are discussing.

Your answer here – keep the asterisks

Finally, answer the following questions:

Discussion Question 1. What are examples of environmental effects that you think affect the sunflower traits measured?

Your answer here – keep the asterisks

Discussion Question 2. Do you think the differences among individuals are due mostly to genetic or environmental effects? Why or why not?

Your answer here – keep the asterisks

Discussion Question 3. There is clear evidence from your data that some individuals are more fit (have greater potential reproductive success) than other individuals. To what extent do you think the observed differences are due to chance? Justify your answer.

Your answer here – keep the asterisks

Checklist

☐ Assembled data set (1 pt)

☐ Visualizations of size and number of seeds per plant (1 pt)

☐ Visualization of geographic distribution of number of seeds per plant (1 pt)

☐ Bivariate plot of height and number of seeds (1 pt)

☐ Pair of histograms showing untransformed and log-transformed data (1 pt)

☐ Calculation of Nt+1 (1 pt)

☐ Calculation of proportion of seeds that do not make it if Nt+1 = Nt (1 pt)

☐ Answers to 9 questions (9 pts)

☐ Make a claim and support claim with evidence (4 pts)

20 points total